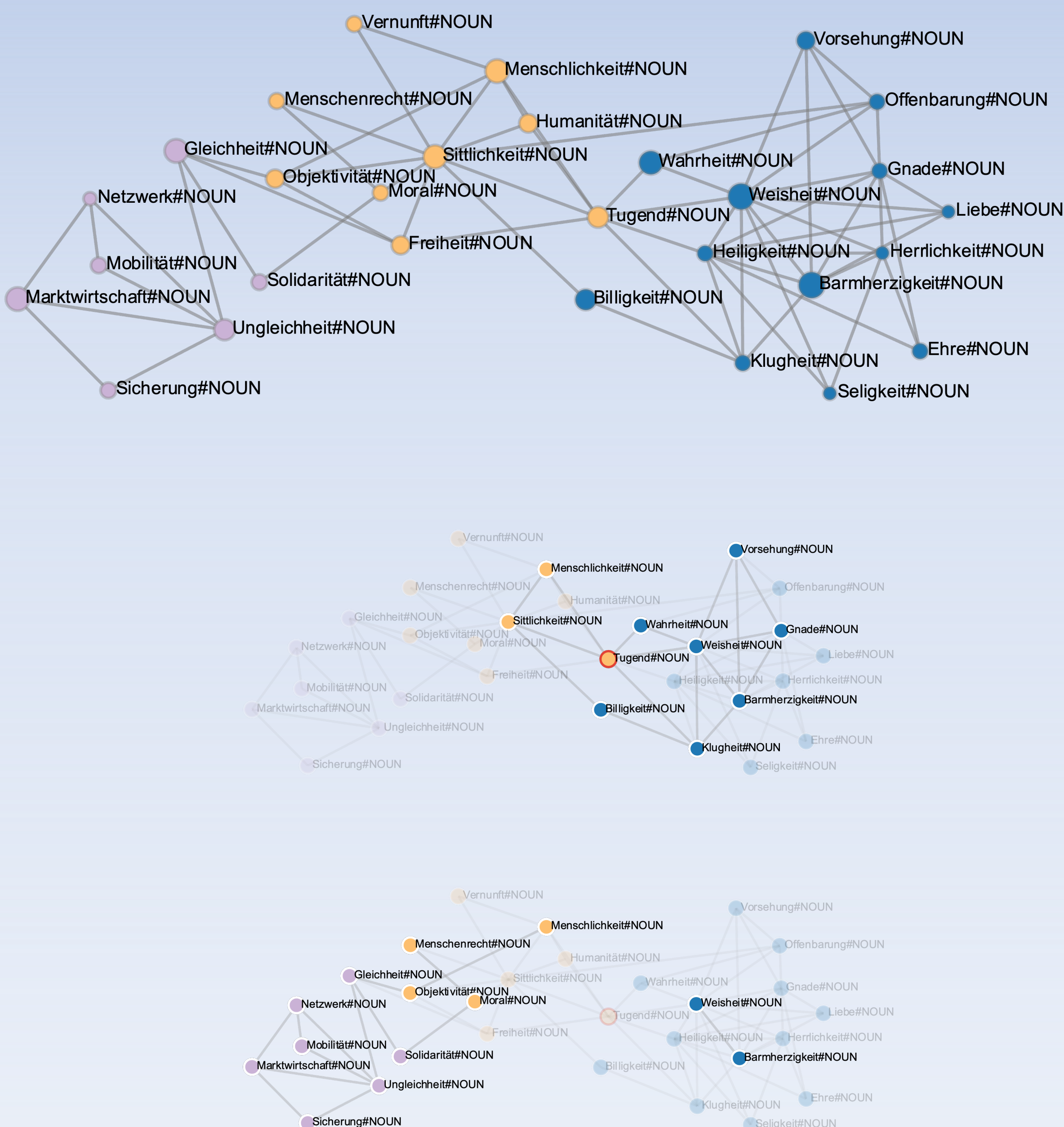


Digitale Begriffsgeschichte

Das Tool *Sense Clustering over Time* (SCoT)

Alexander Friedrich

Institut für Philosophie, TU Darmstadt



SCoT (*Sense Clustering over Time*) ist ein **Open-Access-Online-Tool** zur Erforschung der historischen Semantik von Begriffswörtern.

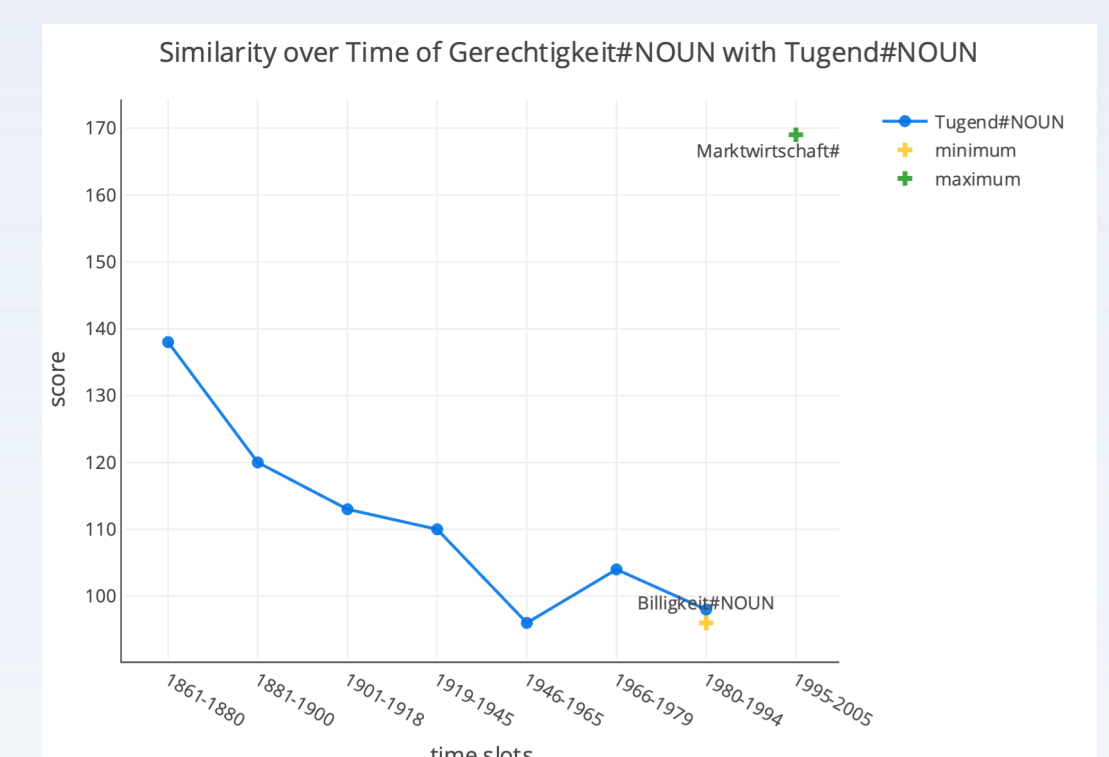
- Basierend auf der statistischen Analyse des Sprachgebrauchs in großen Textkorpora ermittelt SCoT **strukturell-semantische Ähnlichkeitsbeziehungen** zwischen einzelnen Wörtern. Worte werden als Knoten in einem Netzwerk visualisiert. Die Ähnlichkeitsbeziehung dieser Knoten untereinander wird durch Linien repräsentiert.
- Links (oben) sind die 30 ähnlichsten Worte für den Begriff **Gerechtigkeit** in dem deutschen Google Books Korpus im Zeitraum 1861–2019. Je höher die semantische Ähnlichkeit (*similarity score*), desto größer die Knotendarstellung. *Barmherzigkeit* und *Wahrheit* erscheinen hier also *Gerechtigkeit* ähnlicher als *Liebe* oder *Tugend*.

SCoT erlaubt die halbautomatische **Induktion unterschiedlicher Wortbedeutungen (senses)** und deren Veränderung über die Zeit.

- Basierend auf den Ähnlichkeitsbeziehungen der Knoten untereinander gruppiert SCoT die Worte in **verschiedenfarbigen Clustern**. Die Cluster repräsentieren verschiedene Wortbedeutungen (*senses*).
- Links (oben) erscheinen **drei Bedeutungen von Gerechtigkeit**: Das blaue Cluster lässt sich als *religiöser*, das violette als ein *politisch-sozialen* Sinn des Begriffs deuten. In dem gelben Cluster gibt sich der *philosophisch-ethische* Sinn zu erkennen.
- Die **diachrone Analyse** des semantischen Netzwerks zeigt, dass 1901–1918 (links Mitte) der religiöse Sinn noch dominant war, 2006–2012 (links unten) aber zugunsten des politisch-sozialen Sinns verblieben ist. Die *Tugend* (im gelben Cluster) verliert in dieser Zeit ihre Ähnlichkeitsbeziehung zu *Gerechtigkeit* und verblasst.

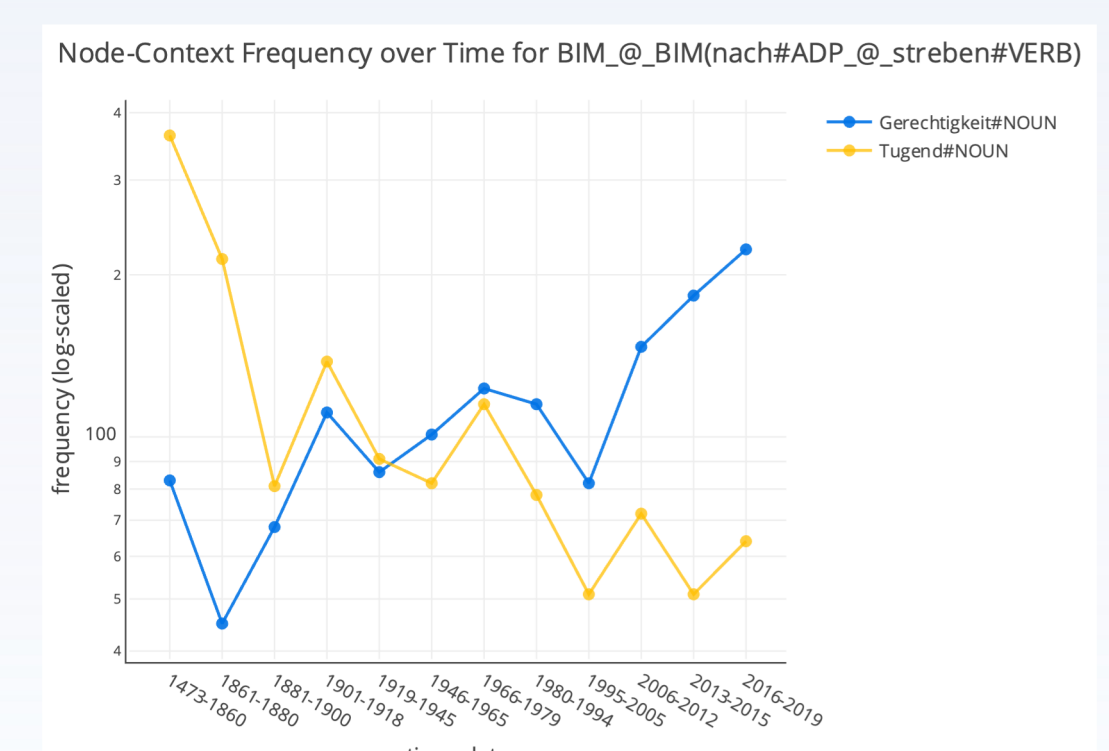
Semantische Ähnlichkeitsbeziehungen lassen sich diachron darstellen.

- SCoT zeigt die jeweiligen Ähnlichkeitswerte (*similarity scores*) pro Zeitabschnitt in einem Graphen und visualisiert so deren Entwicklung über die Zeit.
- So lässt sich etwa sehen, dass der Ähnlichkeitswert von **Gerechtigkeit** und **Tugend** zwischen 1861–1945 deutlich abnimmt und bis heute an Signifikanz verliert (Abbildung rechts oben). Den höchsten Ähnlichkeitswert (*similarity score*) erreicht im gesamten Zeitraum um das Jahr 2000 herum die **Marktwirtschaft**. Diese gehört offenkundig dem dominant werdenden politisch-sozialen Cluster an.



Der Ähnlichkeitswert beruht auf **signifikanten Kontexten**.

- Typische syntaktische Verbindungen**, die zwei Worte miteinander teilen, machen sie einander ähnlich(er). Je größer die Gemeinsamkeit, desto „synonymer“ werden sie.
- Syntaktisch betrachtet ist *Gerechtigkeit* z.B. etwas, wonach *gestrebt* wird. Die Wendung „nach ... streben“ ist auch signifikant für die *Tugend* – seit 1860 aber zusehends seltener (Abbildung rechts unten). Die *Gerechtigkeit* zeigt hier eine gegenläufige Tendenz: „nach **Gerechtigkeit** streben“ ist also eine häufiger werdende Wendung in dem betrachteten Korpus (Google Books). Warum?
- SCoT erlaubt eine **Ausgabe aller relevanter Textstellen** mit dieser Wendung für die weitere Analyse.



SCoT wird in Zusammenarbeit mit der **Language Technology Group, Universität Hamburg** (Leitung: Prof. Dr. Chris Biemann) entwickelt. In dem Verbundprojekt „Das 20. Jahrhundert in Grundbegriffen. Lexikon zur historischen Semantik in Deutschland“ am **Leibniz-Zentrum für Literatur- und Kulturforschung Berlin** (Leitung: PD Dr. Falko Schmieder) sowie in den ERC Synergy Projekt „The Cultures of the Cryosphere. Politics, Infrastructures and Futures of Artificial Cooling“ am **Institut für Philosophie, Technische Universität Darmstadt** (Leitung: Prof. Dr. Alexander Friedrich) wird SCoT erstmals systematisch für begriffsgeschichtliche Forschungszwecke eingesetzt.